DTU Library

Revisiting conceptual stormwater quality models by reconstructing virtual state-variables

Sandoval, Santiago; Bertrand-Krajewski, Jean-Luc; Vezzaro, Luca

Published in:
Proceedings of the 14th IWA/IAHR International Conference on Urban Drainage

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Sandoval, S., Bertrand-Krajewski, J-L., & Vezzaro, L. (2017). Revisiting conceptual stormwater quality models by reconstructing virtual state-variables. In *Proceedings of the 14th IWA/IAHR International Conference on Urban Drainage* (pp. 2316-2323). IWA Publishing.

ICUD-0402 Revisiting conceptual stormwater quality models by reconstructing virtual state-variables

S. Sandoval[1], J. Bertrand-Krajewski[1] L. Vezzaro[2]

1 *Université de Lyon, INSA Lyon, DEEP, EA 7429, 34 avenue des Arts, F-69621 Villeurbanne cedex, France*
2 *DTU ENVIRONMENT, Department of Environmental Engineering, Technical University of Denmark, Building 115, 2800 Kgs. Lyngby, Denmark.*

Summary

This work proposes a Bayesian non-informative reconstruction of virtual state variables in the representation of stormwater TSS pollutographs by two general formulations of the traditional wash-off models, based on 255 rainfall events measured in a 185 Ha French urban catchment. Results revealed the missing representation of an essential process in the traditional Rating Curve model for 56% of the rainfall events. These potential missing processes are hardly interpretable in terms of a unique state of virtual available mass over the catchment that is decreasing over time, due to their high unrepeatability regarding their shape and their low prediction capacity for other rainfall events.

Keywords

conceptual modelling, error models, identifiability, time-variable parameters, existence and unicity of solutions, functional data clustering

Introduction

During the past 40 years, modelling the dynamics of stormwater Total Suspended Solids (TSS) loads at the outlet of urban catchments has been mainly addressed by the idea of accumulation/wash-off (originally by Sartor et al., 1974). Although this conceptual model does not constitute a rigorous physical description of the system, it is usually considered as a physically-based model in the sense that it is based on mass conservation principles and its parameters can be interpretable from a physical point of view (Bonhomme and Petrucci, 2017). Indeed, this accumulation/wash-off model has been implemented in a massive amount of literature and a wide variety of urban catchments (size, land use, complexity), for purposes such as real time control, climate change assessment, risk analysis, water management, and in multiple commercial software tools. However, the unsatisfactory performance of this approach is frequently reported, as well as the difficulty of generalizing its results to real world applications, especially as urban catchments are large and complex (e.g. Vaze et al., 2003; Deletic et al., 2009).

The accumulation/wash-off original model relies on the idea of a virtual "mean" state of TSS mass over the catchment that is accumulated during dry periods and is later on washed off by rainfall and runoff, producing the TSS load at the outlet of the catchment. However, this notion from Sartor et al. (1974), followed by various analogue formulations (e.g. Egodawatta et al., 2007; Freni et al., 2009; Crobedu and Bennis, 2011) has been developed under the following experimental/methodological settings: (i) controlled experimental conditions in small scale systems, (ii) limited number of TSS data, (iii) limited number of rainfall events, and (v) limited assessment of uncertainty in data and model parameters. This suggests how uncertainty linked to application of these models outside their limits (e.g. in large and complex urban catchments – see Liu et al., 2012) is expected to be amplified.

Furthermore, the existence of a "mean" state of available TSS mass for large catchments (from 100-200 ha) has been questioned from a low identifiability and a high spatial variability of models parameters (Bonhomme and Petrucci, 2017). This hypothesis can be nourished from satisfactory results when applying a simplified rating curve RC model (Huber et al., 1988) without including a "mean" available mass term (e.g. Kanso et al., 2005). Under this perspective, one can ask for the existence of a deterministic global process missed by the RC model, essential to represent the pollutant loads, which is oversimplified or misinterpreted by the accumulation/wash-off idea.

For this purpose, the Time Variable Parameters (TVP) concept has been introduced in the hydrological and environmental context as a powerful statistical model-based approach to describe unobserved processes or state variables (e.g. Young, 2013). The TVP method reconstructs the potential time variability of the optimal parameters for a given model. TVP can be used with transfer function models estimated using instrumental variables (e.g. Young, 2013), or with general conceptual models within Bayesian frameworks (Renard et al., 2010). Indeed, Bayesian Total Error Analysis BATEA (Kuczera et al., 2006) has emerged as a promising model-based TVP estimation technique for reconstructing unmeasured inputs or state variables, adaptable to non-linear and complex model structures, including flexibility in the error model of the reconstructed state variable (e.g. Sun and Bertrand-Krajewski, 2013).

This paper aims to revisit the accumulation/wash-off idea by using Bayesian TVP reconstructions, scrutinizing for evidence of a deterministic global process and its interpretability as a "mean" state of available pollutant mass being washed by rainfall. Furthermore, this accumulation/wash-off model structure is sought to be reformulated by analysing the inter-event repeatability of the reconstructed state variables. The experimental/methodological settings aimed at overcoming the limitations of previous studies by considering: (i) an urban catchment of 185 ha, (ii) TSS load measurements with a temporal resolution of 2 minutes (with data uncertainty), (iii) 255 rainfall events (from 2004 to 2011), (iv) the state-of-art calibration and parameter uncertainties assessment (Bayesian methods, e.g. Kuczera et al., 2006).

Materials and Methods

Accumulation/wash-off model

The original accumulation/wash-off model from Sartor *et al.* (1974) describes the stormwater TSS load [kg] at the outlet of an urban catchment as the product of two factors: a sediment mass $M(t)$ [kg] and a removal factor $W(t)$ [-]. In this representation, $M(t)$ is a decaying state variable of a virtual available stock mass which in the end limits the load production given by a wash-off process $W(t)$. Sartor *et al.* (1974) describe the two terms as follows, including a set of calibration parameters:

$$load(t) = M(t) \cdot W(t) = M_0 \cdot \, e^{-a \cdot Q(t)^r \cdot t} \cdot a \cdot Q(t)^r \qquad (1)$$

where M_0 [kg] is the initial condition from which M starts to decay ($M(0) = M_0$, with $t = 0$ the beginning of a rainfall event), and a and r express the influence of the stormwater flow $Q(t)$ [m^3/s] on the removed TSS mass. Several different formulations for M and W can be found in the literature (Egodawatta *et al.*, 2007; Freni *et al.*, 2009). Among these, the simplest one is called "infinite stock" or Rating Curve RC (Huber *et al.*, 1988), where M is considered constant ($M = M_0$) and the wash-off term shows a non-linear relation to the flow:

$$\mathrm{load(t)} = M_0 \cdot Q(t)^r \qquad (2)$$

In this case, the calibration parameters are reduced to two (M_0 and r).

As a generality, it can be shown that for most model structures $0 < M(t) \leq M_0$ and $\mathrm{d}M/\mathrm{d}t \leq 0$ for all t, M being described as either a time variable or a constant process. Indeed, the RC model in Eq 2 can be directly linked with any time variable formulation of M (as in Eq 1), by making M_0 a time variable parameter (TVP). Therefore, this work proposes to undertake Bayesian reconstructions of the "virtual" state variable M by adopting the RC traditional model in Eq 2 and replacing M_0 by a TVP $\widehat{M}(t)$, keeping r as a calibration parameter (formulation F1). The $\widehat{M}(t)$ estimation in F1 can be directly compared to a time constant or variable traditional M formulation (*e.g.* Eq 1). As a complementary analysis, a second formulation (F2) is explored, in which the RC model is modified with $\hat{r}(t)$ [-] as the TVP to be reconstructed and M_0 as a calibration parameter.

Experimental setup

The proposed models are tested with information about 255 rainfall events from the Chassieu urban catchment (Lyon, France), measured between 2004 and 2011. This is one of the experimental sites of the OTHU project (Field Observatory for Urban Hydrology - www.othu.org). The basin is a 185 ha industrial area drained by a separate storm sewer system, with imperviousness and runoff coefficients of about 0.72 and 0.43 respectively. Flow rate Q [L/s] and TSS concentrations [mg/L] at the outlet of the catchment are measured in a 1.6 m circular concrete pipe with a 2 minute time-step resolution. Q is calculated from water depth with a relative standard uncertainty from 15 % to 25 %. Regarding TSS concentrations, the drainage water is pumped into an off-line monitoring flume in a shelter, where turbidity measurements (NTU) are converted into TSS concentration values by local calibration functions (see Sun et al., 2015). The standard uncertainties in the estimated TSS concentrations ranged from 11 % to 30 %. The uncertainties of the TSS load [kg] are propagated by the Law of Propagation of Uncertainties (LPU) (ISO, 2009), obtaining relative standard uncertainties from 15 % to 50 %. Standard uncertainties become higher for higher values of the measurands (Q, TSS concentration, load).

Bayesian reconstruction of virtual state variables by Time Variable Parameters

The reconstruction of a virtual state variable by Time Variable Parameters (TVP) consists in solving a calibration problem where a TVP is represented as an additional time series of parameters. To avoid over-parametrization, the TVP time series has usually a coarser temporal resolution than the output data (from Sun and Bertrand-Krajewski, 2013). This TVP can be estimated jointly with the other set of calibration parameters θ (r for F1 and M_0 for F2) by means of a Bayesian inference scheme. Therefore, θ and *TVP* are defined as random variables, where their joint posterior probability density function (PDF) is calculated by Eq 3 as $p(\theta, TVP/ Q, load)$, given the input $Q(t)$ and output $load_{obs}(t)$ data with some prior knowledge about θ and *TVP* (from BATEA in Kuczera *et al.*, 2006). $p(\theta, TVP/ Q, load)$ is a posterior probabilistic characterization of θ and *TVP* ($\widehat{M}(t)$ for F1 or $\hat{r}(t)$ for F2), in which the values with the maximum likelihood are assumed be the "optimal" parameters θ_{opt} and TVP_{opt} ($\widehat{M}(t)_{opt}$ for F1 or $\hat{r}(t)_{opt}$ for F2).

$$p(\theta, TVP/Q, load)\ \alpha \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}load_t^2}} exp\left[-\frac{1}{2}\frac{(load_{sim}(Q(t),\theta,TVP) - load_{obs}(t))^2}{\hat{\sigma}load_t^2}\right] \cdot P(\theta) \cdot \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}TVP^2}} exp\left[-\frac{1}{2}\frac{Var(TVP)}{\hat{\sigma}TVP^2}\right] P(\mathrm{TVP}) \quad \text{Eq (3)}$$

where n is the total observed load values in $load_{obs}(t)$. $load_{sim}(Q(t)_i, \theta, TVP)$ is the simulated load by the input flow rate series $Q(t)$ and a set of θ and *TVP*. $p(\theta)$ and $p(TVP)$ represents a prior belief about the probability that a candidate set of θ and TVP values are "true" (assumed as a non-informative uniform distribution for all cases).

Equation 3 allows to explicitly separate the model error of *TVP* (second Pi product) from the error model of the output *load* (first Pi product). With the purpose of finding a TVP estimation "as constant as possible" (and therefore less informative), the error model of *TVP* (second Pi product) is assumed to be proportional to TVP's own variance *Var*(*TVP*). Both error models, for the *load* and TVP estimations (first and second Pi product resp.), are assumed to be independent and normally distributed, with the error variances $\hat{\sigma}load_t^2$ and $\hat{\sigma}TVP^2$, respectively. $\hat{\sigma}load_t^2$ is considered heteroscedastic, being equal to the square of the standard uncertainty of each observed value loadobs(t) (e.g. Sun and Bertrand-Krajewski, 2013). On the other hand, $\hat{\sigma}TVP^2$ is considered to be homoscedastic and is estimated as another parameter in the set θ (e.g. Sage et al., 2015), expressed as $\hat{\sigma}M^2$ [kg] for F1 and $\hat{\sigma}r^2$ [-] for F2. The same Eq 3 is used for calibrating the RC model, by omitting the second Pi product term, delivering p(θ/Q, load) as a posterior probabilistic characterization of θ (with θopt also for the "optimal" parameters).

The DREAM algorithm (Vrugt, 2016) is applied to solve Eq 3 and to evaluate the three formulations (the traditional RC, the F1 and F2 formulations) for the different θ and TVP variants (Tab. 1). The logarithmic form of Eq 3 is implemented to ensure numerical stability and a max. number of simulations of 6x105 is used with a max. of 30 parallel Markov Chains to reach convergence (Gelman and Rubin RB convergence criteria > 1.2, see Gelman and Rubin, 1992). The TVPs reconstruction is undertaken with a resolution of 12 time windows (i.e. equivalent to 12 parameters, see Tab. 1), balancing between the convergence of the algorithm (RB > 1.2) and capturing the global dynamics of the TVPs (see application in Sun and Bertrand-Krajewski, 2013). Tab. 1 summarizes the parameters min. and max. values search ranges in the prior uniform distributions p(θ) and p(TVP) in Eq 3. These ranges are defined for r and M0 from the literature (Kanso et al., 2005) and are equally adopted for each window of $\widehat{M}$(t) and $\hat{r}$(t). The max. of the error variances $\hat{\sigma}M^2$ and $\hat{\sigma}r^2$ are defined, respectively, as 4 times the standard deviation of M0 and r in the RC calibration.

***Tab. 1**. DREAM solving of Eq 3 for virtual state variables reconstruction formulations, specifying the set of parameters ϑ and TVP, the total number of parameters and the min. and max. search range in the prior distributions p(ϑ) and p(TVP).*

Formulation	**Parameters θ**	**Parameters in TVP**	**Number of parameters θ + TVP**	**Min/max search range values**
RC model	*r* (-), M_0 (kg)	-	2	*r* [0, 5] ; M_0[0 , 8e5]
F1	*r* (-), $\hat{\sigma}M^2$ (kg)	$\widehat{M}(t)$ (kg)	2 + 12 = 14	*r* [0, 5] ; $\widehat{M}(t)$ [0 , 8e5] for each *t* ; $\hat{\sigma}M^2$ [0 , 2.4e5]
F2	M_0 (kg), $\hat{\sigma}r^2$ (-)	$\hat{r}(t)$ (-)	2 + 12 = 14	M_0 [0 , 8e5] ; $\hat{r}(t)$ [0 , 5] for each *t* ; $\hat{\sigma}r^2$ [0 , 1.5]

The parameters are estimated for each individual rainfall event (event-based calibration). This eliminated the need for a "dry build up" model (e.g. Freni et al., 2009) as the initial sediment mass (M0) is estimated for each event.

For each i-th calibration event, the parameter estimation results (posterior probability and optimal parameter sets) provide the basis for the model evaluation. This is performed by looking at: (i) the Nash-Sutcliffe efficiency (NSi) between simulated (loadsim(t)i) and observed (loadobs(t)i) loads, besides the well-posedness and identifiability of the Bayesian inference p(θ, TVP /Q, load)i (intra-event identifiability); (ii) similarities in the shape or dynamics of TVPopt i with estimations for other events (inter-event identifiability from repeatability); (iii) the capacity of a given set of θopt i and TVPopt i to represent another rainfall event measured by NSi (inter-event transferability, see Bardossy and Singh, 2008) and (iv) formulate further hypotheses about a potential missing process based on physical knowledge about the system and the obtained results (interpretability).

Results and Discussion

The available data and simulation results for event 16 are shown as an example in Fig. 1 and Fig. 2 (event from 23/09/2004 22:00 to 24/09/2004 07:00). The input hydrograph Q(t) is shown in Fig. 1a, while the observed TSS loadobs(t) is shown in Fig. 1b.

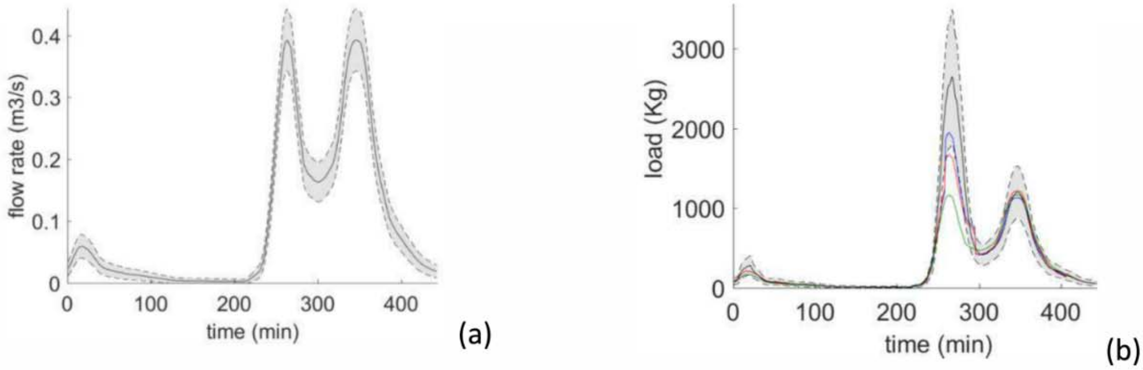

Fig. 1. *a) measured Q hydrograph (solid black) and b) measured TSS load pollutograph (solid black) with 95 % coverage intervals (grey bands) for event 16. Simulation results with optimal parameter sets ϑopt 16 and TVPopt 16 for RC, F1 and F2 are shown as green, blue and red solid lines respectively.*

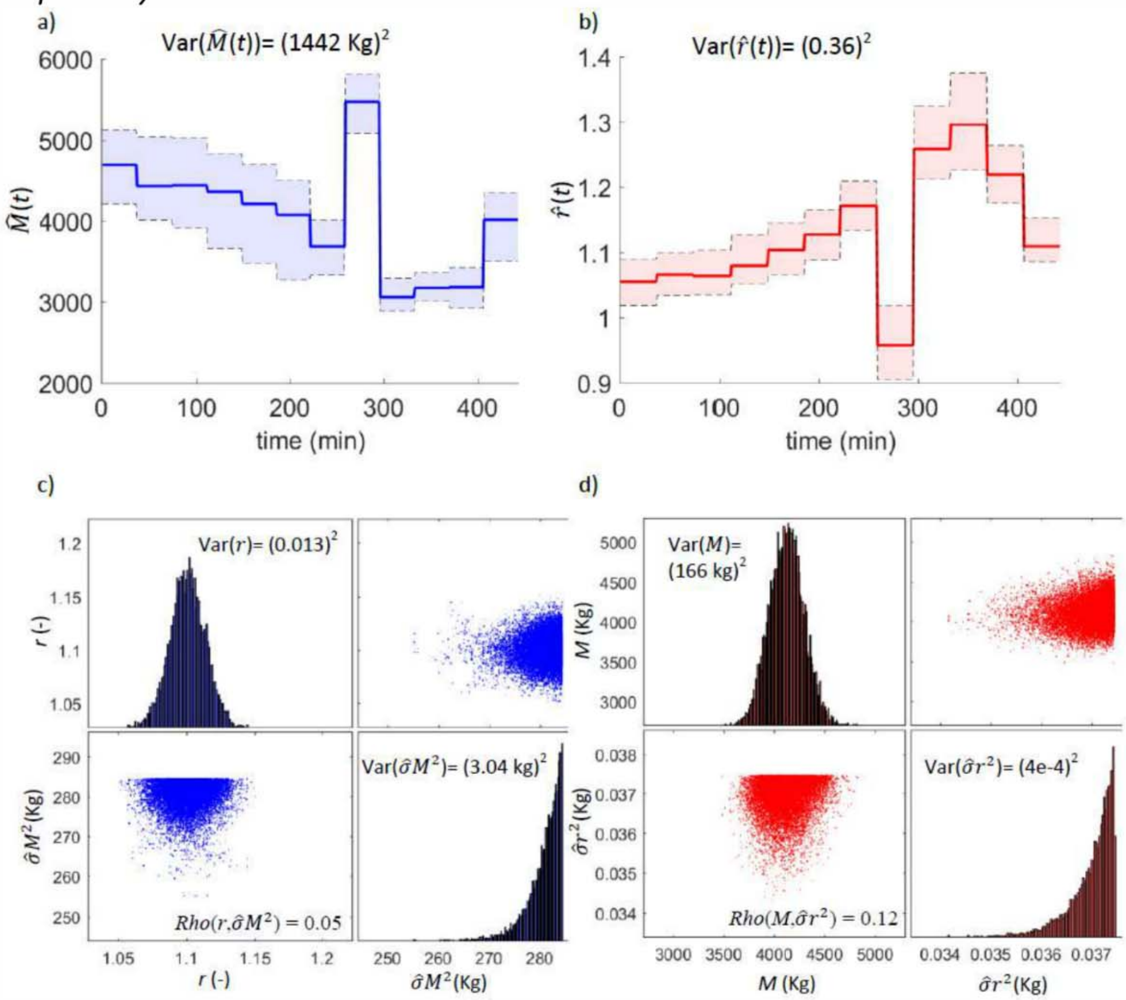

Fig. 2. Results from event 16, *$TVP_{opt\ 16}$ reconstructed time series (solid line) with 95 % coverage intervals (coloured bands). a) $\widehat{\boldsymbol{M}}(t)_{16}$ for F1(blue). b) $\hat{\boldsymbol{r}}(t)_{16}$ for F2 (red).* The correlation matrix of ϑ_{16} *for c) r_{16} [-]; $\widehat{\boldsymbol{\sigma}}\boldsymbol{M}^{2}{}_{16}$ [kg] (F1 blue) and d) $M_{0\ 16}$ [kg]; $\widehat{\boldsymbol{\sigma}}\boldsymbol{r}^{2}{}_{16}$ [-] (F2 red).*

Intra-event Identifiability – Although there is an intrinsic correlation between RC model parameters (therefore for F1 and F2) due to the mathematical structure of the model (Kanso *et al.*, 2005), the intra-event identifiability regarding θ and *TVP* gives promising results. The RC model local calibrations report unsatisfactory adjustments between the simulated and measured loads, with $NS_i < 0.8$ for 142 of the 255 events (56 %). The F1 or F2 formulations achieve greater NS values in all the events, increasing the values of $NS_i > 0.8$ for 60 % of cases in which RC reported $NS_i < 0.8$. In the example shown in Fig. 2, the NS for the RC model (green) is 0.65, while the F1 (blue) and F2 (red) formulations show NS above 0.8. This analysis encourages the applicability of the studied model structures with local estimations of $\theta_{opt\ i}$ and $TVP_{opt\ i}$, supporting the reasoning behind the F1 and F2 reconstructions as potential processes unrepresented by RC.

However, these preliminary results should be analysed under a transferability perspective, as this improvement in NS values can be simply a numerical effect resulting from increasing the number of parameters in F1 or F2 formulations. For the case of $TVP_{opt\ i}$, undesired correlations between the measured $load_{obs}(t)_i$ and the TVP parameters $\widehat{M}(t)_{opt\ i}$ or $\hat{r}(t)_{opt\ i}$ are above 0.6 and 0.5 (respectively) for only 25 % of the events. This result brings evidence that the $TVP_{opt\ i}$ reconstructions are contributing with additional information (also as NS values are higher than for RC), without mimicking the measured *load* dynamic.

Inter-event identifiability from repeatability - A functional clustering by *k*-centre method is applied to identify groups of the TVP_{opt} time-varying curves with similar shapes (Chiou and Li, 2007). A number of *k* groups is defined in order to visualize *k* different potential "repeated" behaviours in the set of optimal TVP_{opt} curves. For interpretability, each of the $TVP_{opt\ i}$ curves is standardized by the transformation $z(TVP_{opt\ i})$ with zero mean and unitary standard deviation (Kreyszig, 1979). For the F1 formulation, applying the Chiou and Li (2007) method with $k = 2$ groups allows separating the $\widehat{M}(t)_{opt}$ into two "similar" clusters, shown in Fig. 3 (light and dark blue groups of curves, corresponding to 57 % and 43 % of the events). The means of each group are shown in the figure on the right, along with the corresponding 95 % coverage intervals.

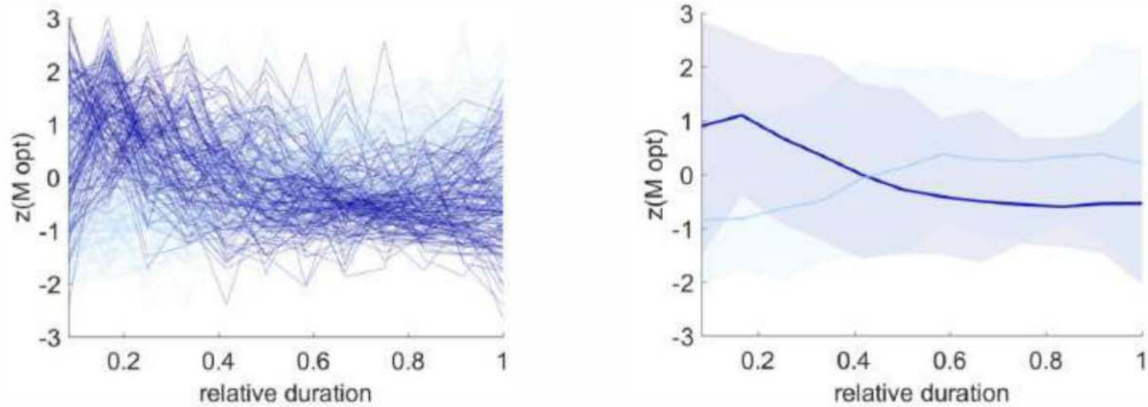

***Fig. 3**. (left) Curves of TVP parameters ($\widehat{M}$(t)opt i) grouped by k = 2 clusters (light and dark colours groups). (right) mean curve (light and dark blue solid lines) and 95 % confidence intervals (light and dark blue bands) for each group.*

The traditional interpretation of M can be preliminarily associated to the dark blue group, as the mean curve shows an apparent decaying trend (Fig. 3). However, this hypothesis would lack of an immediate physical interpretation for the remaining 57 % events (light blue group). Furthermore, the trends of the mean curves for the dark blue group (negative trend) and light blue group (positive trend) show to be statistically insignificant, due to the strong variability of the curves inside each light or dark blue group. This effect of randomness is stronger when the variability of $\widehat{M}(t)_i$ (estimated by calculating the 95% coverage intervals from the posterior distributions) is considered. Similar conclusions are drawn for clustering the $\widehat{M}(t)_{opt\ i}$ curves into more groups ($k > 2$), and for the

F2 formulation. These results reveal the difficulty in identifying or characterizing a unique virtual process potentially missed by the RC regarding an inter-event scale, given the lack of repeatability of the shapes of the $TVP_{opt\,i}$ curves.

Inter-event transferability - The transferability of TVPs is analysed, investigating how a given $TVP_{opt\,i}$ time series is able to reproduce another event from the dataset. Results for $\widehat{M}(t)$ or $\hat{r}(t)$ are analogue, therefore discussion focuses on $\widehat{M}(t)_{opt,i}$ estimations. Each of the 29 most transferable estimations of $\widehat{M}(t)_{opt\,i}$ is able to explain at least 30 rainfall events (NS > 0.8). On the other hand, for the "optimal" local estimations $\theta_{opt\,i}$ of RC, 60 estimations are able to explain at least 30 events (NS > 0.8). The flatter curves from $\widehat{M}(t)_{opt\,i\,=\,1:255}$ tend to be more transferable, as they resemble the constant values in the RC formulation. These results stress the low transferability of the potential missing processes (F1 or F2) to further rainfall events.

Interpretability - These results bring evidence of a potential missing process in the RC model. Although both processes ($\widehat{M}(t)$ or $\hat{r}(t)$) are good candidates to explain RC obstacles from an intra-event analysis, there is no evidence that F1 or F2 is a more valid approach than the other. Indeed, the F1 or F2 formulations show the same explanatory capacity, in the sense that none of them performs better. This highlights the challenge in terms of identifiability and unicity of a potential process missed by the RC model structure, which is hardly identifiable from an inter-event analysis.

Conclusions

This work suggests the missing representation of an essential process in the traditional Rating Curve model based on the observations from 255 rain events. The results indicate that the high unrepeatability of this missing process makes it hardly interpretable in terms of a virtual unique state of available mass in the catchment that is decreasing over time, as assumed by a great number of traditional models. This study shows how high-time resolution quality measurements can provide a support to revisit and question existing models, and to potentially allow developing new stormwater quality model formulations.

Acknowledgements

Authors are grateful to the OTHU project in Lyon, France (www.othu.org) for providing the data set, to COLCIENCIAS and the French-Danish Research Collaboration Program (financed by French Institute in Denmark) for the financial support.

References

Bardossy, A., Singh, S. K. (2008). Robust estimation of hydrological model parameters. *Hydrology and Earth System Sciences*, *12*(6), 1273-1283.

Bonhomme, C., Petrucci, G. (2017). Should we trust build-up/wash-off water quality models at the scale of urban catchments? *Water Research*, 108, 422-431.

Chiou, J.M., Li, P.L. (2007). Functional clustering and identifying substructures of longitudinal data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69, 679-699.

Crobedu, E., Bennis, S. (2011). A new model for simulating TSS washoff in urban areas. Applied and Environmental Soil Science, Article ID 318765, 9 p. doi:10.1155/2011/318765.

Deletic, A., Dotto, C., Fletcher, T.D., McCarthy, D.T., Bertrand-Krajewski, J.L., Rauch, W., Kleidorfer, M., Freni, G., Mannina, G., Tait, S. (2009). Defining uncertainties in modelling of urban drainage systems. In: *8th International Conference on Urban Drainage Modelling*, Tokyo, Japan, pp. 7e11.

Egodawatta, P., Thomas, E., Goonetilleke, A. (2009). Understanding the physical processes of pollutant build-up and wash-off on roof surfaces. *Sci. Total Environ.* 407, 1834e1841. http://dx.doi.org/10.1016/j.scitotenv.2008.12.027

Freni, G., Mannina, G., Viviani, G. (2009). Urban runoff modelling uncertainty: comparison among Bayesian and pseudo-Bayesian methods. Environ. Model. Softw. 24, 1100e1111. http://dx.doi.org/10.1016/j.envsoft.2009.03.003

Gelman, A., Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472.

Huber, W.C., Dickinson, R.E., Barnwell Jr, T.O., Branch, A. (1988). Storm water management model; version 4. *Environmental Protection Agency, United States*.

ISO (2009). ISO/IEC Guide 98-1:2009(E) Uncertainty of measurement – Part 1: *Introduction to the expression of the uncertainty in measurement*. Geneva (Switzerland): ISO, September 2009, 32 p.

Kanso, A., Tassin, B., Chebbo, G. (2005). A benchmark methodology for managing uncertainties in urban runoff quality models. Water Sci. Technol. 51(2), 163–170.

Kreyszig E. (1979). Advanced Engineering Mathematics (Fourth ed.). Wiley. p. 880, eq. 5. ISBN 0-471-02140-7.

Kuczera, G., Kavetski, D., Franks, S., Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, *331*(1), 161-177.

Liu, A., Goonetilleke, A., Egodawatta, P. (2012). Inherent errors in pollutant build-up estimation in considering urban land use as a lumped parameter. J. Environ. Qual. 41, 1690e1694. http://dx.doi.org/10.2134/jeq2011.0419

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research, *46*(5), W05521, 22 p. doi:10.1029/2009WR008328.

Sage, J., Bonhomme, C., Al Ali, S., Gromaire, M.-C. (2015). Performance assessment of a commonly used “accumulation and wash-off” model from long-term continuous road runoff turbidity measurements. Water Res. 78, 47-59. http://dx.doi.org/10.1016/j.watres.2015.03.030

Sartor, J.D., Boyd, G.B., Agardy, F.J. (1974). Water pollution aspects of street surface contaminants. J. Water Pollut. Control Fed. 46, 458-467.

Sun, S., Barraud, S., Castebrunet, H., Aubin, J.B., Marmonier, P. (2015). Long-term stormwater quantity and quality analysis using continuous measurements in a French urban catchment. *Water research*, *85*, 432-442.

Sun, S., Bertrand-Krajewski, J.-L. (2013). Separately accounting for uncertainties in rainfall and runoff: Calibration of event based conceptual hydrological models in small urban catchments using Bayesian method. *Water Resources Research*, 49, 1-14. doi 10.1002/wrcr.20444.

Vaze, J., Chiew, F.H. (2003). Comparative evaluation of urban storm water quality models. *Water Resources Research*, *39*(10).

Vrugt, J.A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273-316.

Young, P. (2013). Hypothetico-inductive data-based mechanistic modeling of hydrological systems. Water Resources Research, 49(2), 915-935.